~~Cooperative Agreement~~
~~University of Montana~~
Thesis/
Reports
Zuuring, H.

FINAL REPORT FOR RESEARCH AGREEMENT
RMRS-98559-RJVA
"An Evaluation of Landscape Metrics and Spatial Statistics for Interpreting SIMPPLE Output"
UNIVERSITY OF MONTANA
FS Contact: Dr. Jimmie Chew
Coop Contact: Dr. Hans Zuuring &
Dr. Mike Sweet

# Final Report

For

# "An Evaluation of Landscape Metrics and Spatial Statistics for Interpreting SIMPPLLE output" (RMRS-98559-RJVA)

Prepared
By

Dr. Hans Zuuring, Mike Sweet,
Charlotte Hilton and Judy Troutwine
School of Forestry
The University of Montana
Missoula, MT 59812
(406-243-6456)

August 23, 2000

# Introduction

This document presents the work completed for a joint research and development project to compute and evaluate various landscape metrics and spatial statistics to determine their utility in conjunction with SIMMPPLE simulations. A progress report was submitted in February 2000 detailing Data Generation Procedures, Software, and Data Analyses procedures for work completed up until January 31, 2000. That report concluded with the intent to "also generate clustered samples from both the Normal and Uniform distributions, run the same factorial experiment, and examine the behavior of the same seven generated metric distributions."

Organization: This document is divided into two parts: one dealing with the procedures for and results from generating clustered and unclustered samples for the three-factor experiment, and another dealing with the evaluation of landscape metrics in conjunction with SIMMPPLE simulations.

## PART I - Landscape metrics behavior

Recall that clustered and unclustered samples were generated for an experiment that was defined by the factors landscape metric, distribution type, and theme class type. As before, **six** landscape metrics were calculated for each distribution type and theme class type. These were:

SHDI = Shannon's Diversity Index
SIDI = Simpson's Diversity Index
MSIDI = Modified Simpson's Diversity Index
SHEI = Shannon's Evenness Index
SIEI = Simpson's Evenness Index
MSIEI = Modified Simpson's Evenness Index

**Two** distributions were considered: **Normal** and **Uniform**. To simulate increasing levels of landscape complexity **seven** theme types were generated having 4, 8, 16, 32, 64, 128, and 256 classes, respectively. These factors yielded 6 x 2 x 7 = 84 treatment combinations and each was replicated for 100 samples, both clustered and unclustered, of size 2,500 observations generated at random.

In order to generate each clustered sample, a portion of the 2,500 observations for the sample was retained as one fixed value. The positions of these observations in the sample resulted from randomly positioned intervals of random length. The remaining positions (between intervals) in each sample of 2,500 observations formed clusters of random variable values, generated in SPSS as before.

# Software

We again utilized Visual FoxPro (a DBMS), Microsoft EXCEL (spreadsheet), SPSS v9.0 (statistical analysis package), Arc/info v7.2.1 (GIS) and an evaluation copy of FRAGSTATS (landscape metrics generator) that is launched from within Arc/Info. Two additional sources were used as well. Arc/Info programs (Arc Macro Language or AML)

were developed to facilitate, enhance, replace, or create new procedures. STAT-GRAPHICS was used to calculate intermediate statistics used in calculating distribution normality test values.

## Data Generation Procedures

Theme Type Coverages

Arc/Info v.7.2.1, SPSS v9.0, and Microsoft Visual FoxPro v.6.0 Excel 97 were used to generate the random samples. Four procedures, first listed and then described below, were required to generate the samples.

1. Generate columns of 0/1 values in 2,500 positions representing random intervals and clusters. Create two sets of 100 columns, one for the Normal distribution and one for the Uniform distribution.

2. Generate random variable values for the distributions of interest (Normal and Uniform).

3. Use the outputs from Procedures 1 and 2 to create clustered samples of 2,500 values. The samples were stored as attribute values for labels in Arc/Info polygon coverages, one for the Normal and one for the Uniform distribution.

4. Generate seven theme types (coverages) based on the number of classes (4, 8, 16, 32, 64, 128, 256) into which the random numbers are categorized.

AMLs were developed to carry out Procedures 1, 3, and 4. Procedure 2 is very similar to that described in the earlier report for nonclustered samples. Note that, although Procedure 4 was earlier performed with FoxPro programs, the generation of clustered samples containing random values in (0,1) (Procedure 3) results in Arc/Info coverages/ tables at the point that the samples are completed. Also, FRAGSTATS requires Arc/Info polygon coverages as input. Thus, Procedure 4 is most conveniently done via Arc/Info capabilities.

**Procedure 1.** Arc/Info capabilities for generating random numbers and writing ASCII files proved to be suited to generating random intervals/clusters of 0/1 values. Each clustered sample of 2,500 values were to be comprised of 40% (1000) fixed values in intervals and 60% (1500) random numbers in clusters of random variable values. Each columns of 0/1 values included 1000 "0" values and 1500 "1" values. The columns were then used in Procedure 3 in conjunction with the files generated by Procedure 2 to generate the clustered samples.

Each interval was to be randomly placed and was to be of random length. In order to use random number techniques to achieve the desired proportion of 0 and 1 values however, the number of intervals could not be fixed or even randomly predetermined.

The solution was to select a random starting position and random length of interval, and then repeat this process until 1500 positions remained outside of the fixed-value intervals. Bounds for the random starting position for each interval were 1 and 1500 inclusive. A condition was set to the effect that, if a starting position occurred within an interval previously created by the program, another random starting point was selected. A new interval could however have a length that overlapped a pre-existing interval. Thus, interval lengths were not determined only by initial, randomly selected lengths. Most intervals in the final state of each 0/1 column were comprised of at least two short intervals. Best results were achieved by setting the bounds for the random selection of an interval length to 1 to 8, inclusive. Final interval lengths for each sample ranged from 1 to over 100 positions.

After the each placement of a new interval, the total number of remaining "1" positions was counted. If the number was > 1500, another interval was generated. When the number remaining approached 1500 + 8, the interval length was calculated by the "number remaining" – 1500. For some 0/1 columns, the "last" interval was repeated once or twice because of overlap with pre-existing intervals. The generation process for each sample ended however with exactly 1000 positions within intervals of "0" values and 1500 positions within clusters of "1" values. These steps are contained in a file named ***cluster01b.aml.***

The advanced user might notice that each set of 100 columns was actually generated in 10 sets of 10 columns. The generating program makes extensive use of arrays stored in variable notation. AML functions produce quicker and more reliable results when array dimensions are kept relatively small. Generating the 0/1 columns in sets of 10 proved to be entirely reliable and take less time than attempting to generate 20 or more columns in one pass.

**Procedure 2.** From the description of Procedure 1, it is clear that, not 2,500, but 1,500 random values from the Normal and Uniform distributions were needed for each sample. Thus, this procedure started as before with a column of sequential numbers created in EXCEL, but in this case the numbers ranged from 1 to 1500 inclusive. The ensuing steps for opening the EXCEL table in SPSS, defining variable names, and computing variable values have been stored in SPSS command syntax files. The two programs are named ***u100rv.sps*** and ***n100rv.sps***.

To run either program, first double click on the filename in Windows explorer. This will start SPSS and open a syntax command window displaying the program. Edit the pathname in the second line of the program to match your directory structure and the location of the table containing the column of 1500 sequential numbers. From the **Run** menu in the syntax window, click on **All** to run the program. When completed, save the file in dBase IV format as ***u1500_100rv.dbf***, for the uniform distribution for example. Any open viewer windows are not needed at this point and may be closed.

The reader may recall that the resulting values for both distributions were to fall between 0 and 1 inclusive and that the values generated by SPSS for the normal distribution needed to be transformed. The FoxPro program, ***trans_norm.prg***,

a slight modification of ***normcalc.prg***, was used to transform the normal distribution random variable values. ***Trans_norm.prg*** includes commands to set a default directory for accessing and saving tables, and to automatically save the output to a dBase IV format. (Although FoxPro reads in the dBase IV format automatically, tables must be exported interactively or copied programmatically with the "fox2x" option to save them as dBase IV format.) The dBase IV format is required for conversion by Arc/Info into its own INFO format (as done in the next procedure).

To run the program, first open the ***trans_norm.prg*** file in Visual FoxPro and edit the path name to match that of the location of your ***n1500_100rv.dbf*** table. Save the file while keeping the program window open. Then click on the "**!**" toolbar button. When done, just close FoxPro. The transformed values file will be stored in the same directory as the input file.

Because the transformation involved only subtracting and dividing by constants (minimum and maximum values), the resulting random variables remain normally distributed (Sokal and Rohlf, pp. 54, 145). The resulting Normal and Uniform random variable tables may be found in ***n1500_100rvt.dbf*** and ***u1500_100rv.dbf***.

**Procedure 3.** This procedure utilizes the outputs from Procedures 1 and 2 to create tables containing the clustered samples. The tables are then joined to the polygon attribute table (.pat) for Normal and Uniform distributions, respectively. The steps to carry out this procedure, performed by ***clustercov.aml*** are as follows. For each (uniform, normal) distribution:

a. Copy the pre-existing coverage, cov50, containing the correct range of 'user-id' values to a name referring to the distribution (norm, unif). Place it in the tempcovs directory for the distribution.

b. Copy the ASCII files containing lists of 0/1 columns to the 'info' directory within the directory tempcovs for the distribution.

c. Since the 0/1 columns were generated in sets of 10 columns, define 10 tables having 10 columns each and load the 0/1 values into the tables.

d. Add 10 fields to each of the 10 tables to hold the clustered random variable values for the 100 samples.

e. Convert the dBase IV file containing the 1500 random variable values to INFO format.

f. Process one of the 10 tables at a time, looping through 10 samples for each. Define two AML cursors, one for a 0/1 table and one for the table containing random variable values. For each sample: Step through the 0/1 values, retrieve the next random variable value when a "1" is encountered, and place the random variable value into the record in the clustered random variable field corresponding to the record of the "1" in the 0/1 field. Where a "0" is encountered, set the value in the clustered random variable field to -99.9999 (to match the F8.4 field format).

The fields containing the 0/1 values may be dropped after the cluster process is completed.

g. Join the 10 tables to the polygon coverage. There should now be 100+ fields in the .PAT, each containing alternating intervals and clusters of random values in the continuous interval (0,1).

h. For each of the 100 fields holding clustered sample values, select the number of records having random variables. Type the result to the watch file for checking.

The program includes a loop command which allows the user to process samples for either one or both samples in one run.

**Procedure 4.** It remains to generate seven coverages corresponding to the seven theme types consisting of 4, 8, 16, 32, 64, 128, and 256 class values, respectively. The algorithm used to calculate the class values was adapted to AML from the FoxPro programs used for the same purpose. There is a difference in the calculations, in that the fixed-value positions in each sample are assigned to class 1 for each theme type. The remaining continuous random values are assigned to 3, 5, 7, 15,..., 255 classes or n-1 classes rather than n classes. The AML, ***classes.aml***, first copies the norm or unif coverage to a new directory, n4 or u256 for example, and names the copied coverage to norm4, norm8, ..., unif256. The program then adds fields to hold the class values, calculates the class values, and removes the fields that are no longer needed. There are seven theme type coverages for each distribution for a total of fourteen theme type coverages.

## DATA ANALYSES

Calculating landscape metrics for theme type coverages

A new version of FRAGSTATS was made available in early February 2000. For our purposes, the method of operation was very similar to that for the previous version. The differences were that each batch run needed to be made in a separate directory and the number of attributes that could be processed in one run was reduced. The first difference required that a coverage having 100 attributes to be processed needed to be copied into seven to ten directories so that 10-13 attributes could be processed in each directory.

Because these differences increased the time and effort to processing attributes, the vendor was notified about the problem. Four coverages (400 attributes) of fourteen were nevertheless processed in this manner. For no apparent reason however, the number of attributes that could be processed in one directory declined further to three to four attributes. This necessitated a more timely solution.

Formulas for landscape metrics calculated by FRAGSTATS are provided in Appendix 3 of *Fragstats: Spatial Pattern Analysis Program for Quantifying Landscape Structure*

(McGarigal and Marks, 1994). Based on the provided formulas, an AML was written to directly calculate the seven landscape metrics for the remaining ten coverages. This program, ***indices.aml*** places the indices' values into INFO tables; these are then exported to dBaseIV format. The program also includes loop commands to allow a user to calculate the metrics for all fourteen coverages in one run.

Calculating landscape metrics for simulation outputs

Once simulation outputs were available in ASCII format, AMLs were used to join data fields to the landscape coverage and to calculate landscape metrics. The program, ***importsim2.aml*** defined four INFO tables to hold values from 25 simulation outputs each for a total of 100 simulations, and then imported the comma delimited values into the tables. Dividing the fields among four tables was necessary to avoid the record length limit of 4096 bytes in INFO that would otherwise occur during subsequent processing. To maintain distinct field names for each simulation, the sample number was appended to the end of field names. For example, the field names were SLINK_1, SPECI_1, SIZE_1, CANOP_1, PROC_1, and TREAT_1 for sample 1. Except for SLINK AND CANOP fields (integer fields), all fields were defined as character fields. After checking the contents of the tables, they were joined to four copies of the landscape coverage.

Landscape metrics were calculated via a modification of ***indices.aml***, called ***indices_s.aml***. In addition to calculating indices for species, size, canopy, process, and treatment, indices were calculated for a "state" field. This field was created by concatenating the characters that represented the values of the species, size, and canopy fields. This produced the largest set of unique values. ***Indices_s.aml*** includes this operation as well as being specific to the other fields of the landscape coverage.

To see how landscape indices vary over time, indices' values were calculated for one simulation for 0 to 15 decades. The process, similar to that for decade 10, was carried out by the programs ***importdecs.aml*** and ***indices_st.aml***.

Evaluating landscape metrics

A criterion for interpreting the landscape metrics was that the indices' values calculated for the normal distribution also be representative of normal distributions. Each set of 100 index values for each metric for each theme type was tested. Both graphs and normality tests were used in the evaluation of the distributions.

The graphs included histograms, box plots, and standard normality plots. The statistical package SAS was used to draw these. To quantify the closeness of each parent distribution to a normal distribution, the D'Agostino-Pearson $K^2$ test for normality (Zar, p. 89) proved to be most appropriate. This test is suitable for discrete data and does not require grouping values into arbitrary categories as does the Chi-square Goodness of Fit test. Also, the Chi-square test is considered to have low power in rejecting false null hypotheses (Zar, p. 87). The D'Agostino-Pearson test does require calculating z-scores for skewness and kurtosis and the sum of the squares of the Z-scores. The resulting

values ($K^2$) are compared to critical values for a Chi-square distribution with 2 degrees of freedom.

Calculation of the Z-scores was done with the STATGRAPHICS program. These scores were then entered into an INFO table. The $K^2$ values and significance values were calculated and assigned interactively in INFO.

Throughout this evaluation, care was taken to ensure that indices' values provided sufficient detail to yield valid results. Sokal and Rohlf (pp. 13-15) recommend retaining enough significant digits to include a range of 50 to 300 possible values. Also, it is desirable to retain digits to facilitate accurately categorized histograms (Sokal and Rohlf, p. 27). Because the landscape metrics do yield a finite number of values (discrete distribution), it was found that five significant digits to the right of the decimal place were required to accurately summarize the indices' values. Further, retaining five digits in both the diversity indices' values and the denominator values for the evenness indices resulted in more accurate evenness indices. For coverage attributes having 256 classes the difference in some evenness indices' values was greater than 4% as compared to retaining only three significant decimal digits.

As a result of these findings, the same procedures were used to test sample distributions of landscape metrics for unclustered samples as well as clustered samples. Graphs and normality test values are included in Word documents, ***n4graphs.doc***, ***n8graphs.doc***, ..., ***n256graphs.doc*** in the respective theme type coverage directories.

## RESULTS

Evaluation of generated metric distributions:

Unclustered samples from a Normal Distribution

D'Agostino-Pearson normality test ($K^2$) indicated that the distribution of each of the six landscape metrics computed in association with each of six thematic classes did not deviate from a Normal distribution, i.e. the null hypothesis of normality was not rejected. These results are based on 100 unclustered samples, each having 2,500 observations drawn from a Normal distribution (Table 1). $K^2$ values tended to increase with increasing number of thematic classes regardless of landscape metric with the exception of SHEI.

Table 1. D'Agostino-Pearson normality test ($K^2$)

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | 1.94 | 2.00 | 1.91 | 1.90 | 1.88 | 2.00 |
| SIDI | 1.83 | 2.50 | 2.37 | 2.38 | 2.58 | 2.84 |
| MSIDI | 1.40 | 2.10 | 1.98 | 2.02 | 2.40 | 2.96 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SHEI | 2.34 | 2.00 | 1.91 | 1.90 | 1.88 | 2.00 |
| SIEI | 1.83 | 2.49 | 2.37 | 2.38 | 2.54 | 2.85 |
| MSIEI | 1.40 | 2.10 | 1.98 | 2.02 | 2.40 | 2.96 |

The mean values for each of the six landscape metrics increased as the number of thematic classes increased. SHDI and MSIDI values as well as SIDI and SIEI values appear to be highly correlated (Table 2). At 256 classes both SHDI and MSIDI approach a value close to 5 while all the other metrics approach a value close to 1.

Table 2. Summary statistics - means

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | .980 | 2.26 | 2.95 | 3.63 | 4.31 | 4.98 |
| SIDI | .574 | .879 | .939 | .969 | .984 | .992 |
| MSIDI | .854 | 2.11 | 2.80 | 3.48 | 4.16 | 4.83 |
| SHEI | .707 | .816 | .850 | .873 | .889 | .898 |
| SIEI | .765 | .938 | .969 | .985 | .992 | .996 |
| MSIEI | .616 | .763 | .808 | .838 | .858 | .871 |

Standard deviations associated with SHDI and MSIDI values appear to be highly correlated and increase as the number of thematic classes increase while the standard deviations associated with the remaining landscape metrics decrease with increasing number of thematic classes (Table 3).

Table 3. Summary statistics - standard deviations

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | .046 | .057 | .057 | .058 | .057 | .056 |
| SIDI | .018 | .007 | .004 | .002 | .001 | .000 |
| MSIDI | .043 | .058 | .058 | .058 | .057 | .056 |
| SHEI | .033 | .020 | .014 | .014 | .012 | .010 |
| SIEI | .025 | .008 | .002 | .002 | .001 | .000 |
| MSIEI | .031 | .021 | .014 | .014 | .012 | .010 |

Although all the skewness and kurtosis values are negative and indicate some departure from normality (both should be around zero if the data come from a normal distribution) histograms and log-normal probability plots do not show this departure to be excessive regardless of landscape metric or number of thematic classes (Tables 4 and 5).

Table 4. Summary statistics - skewness

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | -.296 | -.369 | -.374 | -.367 | -.356 | -.358 |
| SIDI | -.410 | -.491 | -.474 | -.457 | -.447 | -.420 |
| MSIDI | -.313 | -.368 | -.350 | -.335 | -.330 | -.313 |
| SHEI | -.296 | -.369 | -.374 | -.367 | -.356 | -.358 |
| SIEI | -.410 | -.491 | -.474 | -.457 | -.446 | -.422 |
| MSIEI | -.313 | -.368 | -.350 | -.335 | -.330 | -.313 |

Table 5. Summary statistics - kurtosis

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | -.464 | -.409 | -.388 | -.395 | -.403 | -.422 |
| SIDI | -.316 | -.339 | -.341 | -.371 | -.421 | -.486 |
| MSIDI | -.355 | -.427 | -.426 | -.446 | -.499 | -.566 |
| SHEI | -.464 | -.409 | -.388 | -.395 | -.403 | -.422 |
| SIEI | -.316 | -.340 | -.341 | -.371 | -.416 | -.485 |
| MSIEI | -.355 | -.427 | -.426 | -.446 | -.500 | -.566 |

Unclustered samples from Uniform Distribution

D'Agostino-Pearson normality test ($K^2$) was not run for this set because we are dealing with data drawn from a uniform distribution. As before SHDI and MSIDI appear to be highly correlated and increase with increasing number of thematic classes. SIEI remains at a value of 1 over all thematic classes while SIDI approaches a value of 1 (it is a probability really!) at 256 classes. Both SHEI and MSIEI decrease very slightly as the number of thematic classes increases beyond 32 (Table 6). At 256 classes both SHDI and MSIDI approach a value close to 5.5

Table 6. Summary statistics - means

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | 1.39 | 2.77 | 3.46 | 4.15 | 4.83 | 5.49 |
| SIDI | 0.75 | .937 | .968 | .984 | .992 | .996 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MSIDI | 1.38 | 2.77 | 3.45 | 4.13 | 4.80 | 5.45 |
| SHEI | 1.00 | 1.00 | 1.00 | .997 | .995 | .991 |
| SIEI | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| MSIEI | 1.00 | 1.00 | 1.00 | .994 | .990 | .982 |

Standard deviations associated with SHDI and MSIDI and SHEI and MSIEI appear to be highly correlated (MSIDI and MSIEI standard deviations are about double those associated with SHDI and SHEI respectively) and increase as the number of thematic classes increase. Standard deviations associated with remaining landscape metrics decrease with increasing number of thematic classes and approach zero (Table 7). Note that all standard deviations are about 10 times smaller than those associated with landscape metrics that were computed from unclustered samples drawn from a normal distribution.

Table 7. Summary statistics - standard deviations

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | .0004 | .0011 | .0018 | .0024 | .0038 | .0047 |
| SIDI | .0002 | .0001 | .0001 | .0001 | .0001 | .0000 |
| MSIDI | .0007 | .0022 | .0035 | .0047 | .0072 | .0085 |
| SHEI | .0003 | .0004 | .0005 | .0006 | .0008 | .0008 |
| SIEI | .0002 | .0002 | .0001 | .0001 | .0001 | .0000 |
| MSIEI | .0005 | .0008 | .0010 | .0011 | .0015 | .0015 |

As might be suspected both the skewness and kurtosis values indicate severe non-normality for all six landscape metrics at 4 thematic classes. However as the number of classes increase to 256 this non-normality disappears for a landscape metrics (both skewness and kurtosis values hover around zero (Tables 8 and 9).

Table 8. Summary statistics - skewness

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | -1.581 | -.497 | -.302 | -.262 | -.296 | .017 |
| SIDI | -1.558 | -.507 | -.307 | -.323 | -.382 | -.07 |
| MSIDI | -1.563 | -.500 | -.291 | -.293 | -.324 | .034 |
| SHEI | -1.574 | -.495 | -.303 | -.261 | -.299 | .015 |
| SIEI | -1.566 | -.517 | -.293 | -.327 | -.379 | .033 |
| MSIEI | -1.561 | -.498 | -.291 | -.292 | -.323 | .034 |

Table 9. Summary statistics - kurtosis

| Landscape Metric | # of classes | | | | | |
|---|---|---|---|---|---|---|
| | 4 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | |
| SHDI | 3.296 | -.342 | .160 | -.246 | -.010 | -.157 |
| SIDI | 3.204 | -.338 | .144 | -.078 | .091 | .003 |
| MSIDI | 3.212 | -.350 | .093 | -.103 | .036 | -.076 |
| SHEI | 3.252 | -.351 | .156 | -.246 | -.013 | -.156 |
| SIEI | 3.210 | -.315 | .106 | -.079 | .061 | -.155 |
| MSIEI | 3.203 | -.347 | .098 | -.100 | .034 | -.076 |

Clustered samples from Normal Distribution

D'Agostino-Pearson normality test ($K^2$) indicated that the distribution of each of the six landscape metrics computed in association with each of six thematic classes did not deviate from a Normal distribution, i.e. the null hypothesis of normality was not rejected. These results are based on 100 clustered samples, each having 1,500 observations drawn from a Normal distribution (Table 10). Note that 2,500 observations were initially drawn but 1,000 of these observations were set to zero at random intervals to create gaps.

Table 10. D'Agostino-Pearson normality test ($K^2$)

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | 0.88 | 0.38 | 0.59 | 0.54 | 0.31 | 0.23 | 0.31 |
| SIDI | 0.43 | 0.81 | 1.46 | 2.09 | 1.16 | 0.77 | 1.06 |
| MSIDI | 0.15 | 0.40 | 1.03 | 1.78 | 1.02 | 0.75 | 0.97 |
| SHEI | 0.88 | 0.38 | 0.59 | 0.54 | 0.31 | 0.23 | 0.31 |
| SIEI | 0.43 | 0.82 | 1.46 | 1.36 | 1.14 | 0.79 | 0.97 |
| MSIEI | 0.15 | 0.40 | 1.03 | 1.78 | 1.03 | 0.78 | 1.02 |

$K^2$ values tended to oscillate across increasing number of thematic classes for a given landscape metric. This behavior was not observed for $K^2$ values associated with unclustered samples.

The mean values for SHDI, SIDI, and MSIDI increased as the number of thematic classes increased. SIEI values remained about the same across increasing number of thematic classes while SHEI and MSIEI values decreased with increasing number of thematic classes (Table 11).

Table 11. Summary statistics - means

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | 1.14 | 1.58 | 2.01 | 2.44 | 2.86 | 3.27 | 3.66 |
| SIDI | .638 | .747 | .795 | .818 | .829 | .834 | .837 |
| MSIDI | 1.02 | 1.37 | 1.59 | 1.70 | 1.77 | 1.80 | 1.81 |
| SHEI | .820 | .757 | .727 | .705 | .688 | .674 | .660 |
| SIEI | .851 | .854 | .848 | .844 | .842 | .841 | .840 |
| MSIEI | .734 | .661 | .572 | .492 | .425 | .371 | .327 |

Standard deviations associated with SHDI values increased slightly as the number of thematic classes increased while the standard deviations associated with the remaining landscape metrics decreased with increasing number of thematic classes (Table 12).

Table 12. Summary statistics - standard deviations

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | .036 | .039 | .041 | .042 | .041 | .041 | .040 |
| SIDI | .014 | .006 | .003 | .002 | .001 | .000 | .000 |
| MSIDI | .040 | .026 | .016 | .009 | .005 | .002 | .001 |
| SHEI | .026 | .019 | .015 | .012 | .010 | .008 | .007 |
| SIEI | .019 | .008 | .004 | .002 | .001 | .000 | .000 |
| MSIEI | .029 | .012 | .006 | .003 | .001 | .000 | .000 |

Both skewness and kurtosis values associated with all landscape metrics tended to oscillate with increasing number of thematic classes (Tables 13 and 14). Skewness values ranged from -.376 to -.110 and indicated that the metrics were skewed to the left. Kurtosis values ranged from -.168 to +.389. Based on these statistics and displays of histograms and log-normal probability plots I judge all these landscape metrics to be normally distributed.

Table 13. Summary statistics - skewness

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | -.265 | -.204 | -.251 | -.240 | -.182 | -.159 | -.181 |
| SIDI | -.218 | -.274 | -.336 | -.376 | -.285 | -.255 | -.285 |
| MSIDI | -.110 | -.199 | -.286 | -.348 | -.269 | -.250 | -.277 |
| SHEI | -.265 | -.204 | -.251 | -.240 | -.182 | -.159 | -.181 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| SIEI | -.218 | -.275 | -.336 | -.376 | -.285 | -.256 | -.281 |
| MSIEI | -.110 | -.199 | -.286 | -.349 | -.270 | -.253 | -.281 |

Table 14. Summary statistics - kurtosis

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | .154 | -.094 | -.004 | -.008 | -.044 | -.106 | -.125 |
| SIDI | -.071 | .092 | .250 | .388 | .237 | .122 | .194 |
| MSIDI | -.168 | .012 | .179 | .342 | .212 | .124 | .173 |
| SHEI | .154 | -.094 | -.004 | -.008 | -.045 | -.106 | -.125 |
| SIEI | -.071 | .093 | .250 | .389 | .231 | .127 | .162 |
| MSIEI | -.168 | .012 | .178 | .343 | .215 | .130 | .190 |

Clustered samples from Uniform Distribution

Again D'Agostino-Pearson normality test ($K^2$) was not run for this set because we are dealing with data drawn from a Uniform distribution. SHDI, SIDI and MSIDI values increased with increasing number of thematic classes although this increase was less pronounced with SIDI. All the remaining landscape metrics decreased in value with increasing number of thematic classes but MSIEI values decreased at a greater rate and more dramatically from .918 to .329 (Table 15). At 256 classes SHDI was more than double the value of MSIDI. Standard deviations associated with SHDI increased substantially with increasing number of thematic classes but only slightly with SHEI. Standard deviations associated with all other metrics decreased with increasing number of thematic classes.

Table 15. Summary statistics - means

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | 1.33 | 1.84 | 2.30 | 2.73 | 3.15 | 3.55 | 3.94 |
| SIDI | .720 | .788 | .816 | .828 | .834 | .837 | .838 |
| MSIDI | 1.27 | 1.55 | 1.69 | 1.76 | 1.80 | 1.81 | 1.82 |
| SHEI | .961 | .884 | .828 | .787 | .756 | .732 | .711 |
| SIEI | .960 | .901 | .870 | .855 | .847 | .844 | .842 |
| MSIEI | .918 | .747 | .610 | .508 | .432 | .374 | .329 |

Table 16. Summary statistics - standard deviations

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | .0004 | .0007 | .0010 | .0014 | .0020 | .0031 | .0045 |
| SIDI | .0002 | .0001 | .0001 | .0001 | .0000 | .0000 | .0000 |
| MSIDI | .0005 | .0005 | .0005 | .0003 | .0002 | .0002 | .0001 |
| SHEI | .0003 | .0003 | .0004 | .0004 | .0005 | .0006 | .0008 |
| SIEI | .0002 | .0001 | .0001 | .0001 | .0000 | .0000 | .0000 |
| MSIEI | .0004 | .0003 | .0002 | .0001 | .0001 | .0000 | .0000 |

Both skewness and kurtosis values associated with all landscape metrics tended to decrease and approach zero with increasing number of thematic classes up to 128 and then became more negative for 256 classes (Tables 13 and 14). Skewness values ranged from -1.601 to -.035 and indicated that the metrics were skewed to the left. Kurtosis values ranged from -.467 to +2.94. Based on these statistics and displays of histograms and log-normal probability plots I judge all these landscape metrics to be non-normally distributed.

Table 17. Summary statistics - skewness

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | -1.589 | -.927 | -.901 | -.357 | -.354 | -.036 | -.466 |
| SIDI | -1.601 | -.953 | -.849 | -.425 | -.440 | -.112 | -.537 |
| MSIDI | -1.599 | -.943 | -.848 | -.394 | -.445 | -.152 | -.523 |
| SHEI | -1.603 | -.925 | -.897 | -.356 | -.352 | -.035 | -.465 |
| SIEI | -1.581 | -.936 | -.866 | -.394 | -.410 | -.114 | -.560 |
| MSIEI | -1.593 | -.948 | -.862 | -.406 | -.437 | -.167 | -.419 |

Table 18. Summary statistics - kurtosis

| Landscape Metric | # of classes | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| | | | | | | | |
| SHDI | 2.888 | .375 | 1.511 | -.467 | -.301 | -.282 | 1.240 |
| SIDI | 2.931 | .455 | 1.290 | -.353 | -.223 | .184 | 1.312 |
| MSIDI | 2.930 | .441 | 1.296 | -.381 | -.209 | .050 | 1.276 |
| SHEI | 2.940 | .382 | 1.519 | -.467 | -.304 | -.278 | 1.237 |
| SIEI | 2.834 | .427 | 1.406 | -.380 | -.300 | .054 | 1.231 |
| MSIEI | 2.907 | .446 | 1.352 | -.377 | -.204 | .125 | .854 |

# PART II - Landscape metric evaluations as related to multiple SIMMPPLE simulations

## METHODS

The behavior of a number of landscape metrics was investigated as related to five stand characteristics: canopy cover, size, species, state, and process. The geospatial database for the Stevensville West Central area was utilized to demonstrate the behavior of selected landscape metrics as related to multiple simulations of SIMMPPLE. This database was used extensively in the development and demonstration of SIMPPLLE for landscape-level interpretations, and was provided by the Rocky Mountain Research Station.

The SIMMPPLE outputs generated as a result of utilizing various landscape metrics exhibited the properties of the normal distribution. Therefore, interpretations of multiple stochastic simulations of SIMPPLLE can be made using the rich set of descriptive statistics associated with normal distributions. The Stevensville West Central landscape was simulated in SIMPPLLE one hundred (100) times for fifteen (15) decades. The $10^{th}$ decade (100 years) was selected for the demonstration application because we felt it was a sufficiently long enough simulation to provide an acceptable distribution of outcomes. Simulations from the $1^{st}$ and $2^{nd}$ decade would presumably have a narrower distribution than the $10^{th}$ decade.

Landscape metrics were calculated for the results from the $10^{th}$ decade of each SIMPPLLE simulation for the Stevensville West Central landscape. McGarigal and Marks (1994) describe three diversity and three evenness indices that are calculated by FRAGSTATS. **Richness** (number of patch types) and **evenness** (distribution of area among patch types), also known as the compositional and structural aspects of diversity, affect the magnitude of diversity indices. The magnitude of the Shannon diversity index (SHDI) is affected more by richness than by evenness. In fact, rare patch types affect the magnitude of this index more than other indices. In contrast, Simpson's index (SIDI) is affected more by a preponderance of common patch types. Because both are diversity indices, their values usually represent the composition of landscapes.

SHDI is the most popular index, but the magnitude in itself is not considered meaningful. Usually, the values for different landscapes or times are compared. SIDI is an actual probability (that any two randomly selected patches are different) and can be interpreted both directly and by comparison to SIDI values for other landscapes. FRAGSTATS also calculates a modified version of SIDI (MSIDI). A transformation is applied to MSIDI so that it can be interpreted in the same way as SHDI.

For illustration purposes, the Shannon diversity index (SHDI) was selected simply because it does emphasize richness rather than evenness. The selection of a particular metric depends on what characteristics of the landscape are most important to interpretation. The form of the presentation for SHDI would be equally valid for the other landscape metrics discussed in this report. On the following pages, four figures are presented that relate to the following stand characteristics: canopy cover, size, species, state, and process. The first figure is a histogram of all values from the

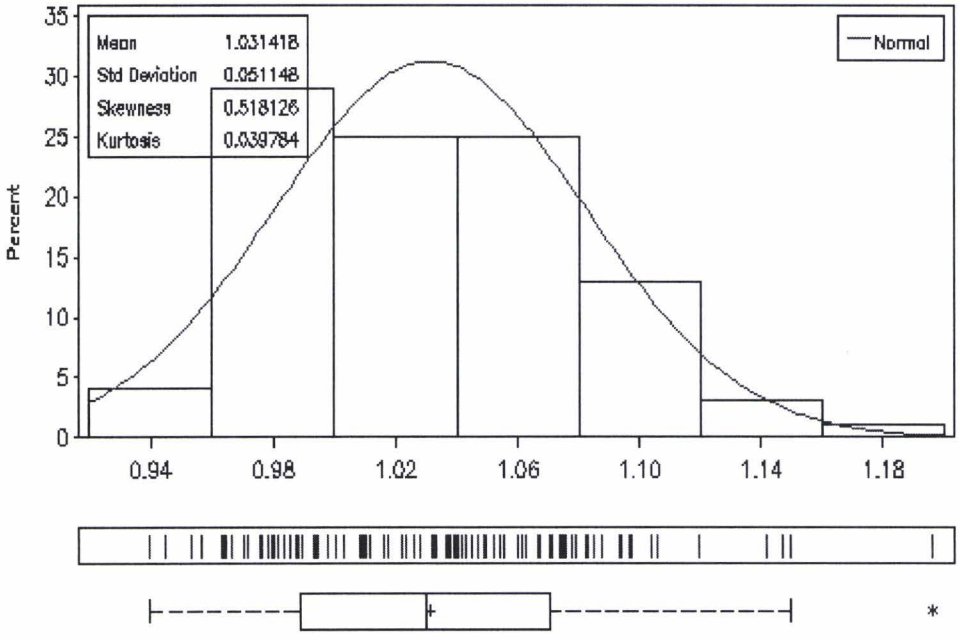
simulations. The histogram is followed by three graphics of a landscape, one depicting the mean of the distribution and the remaining two near the negative and positive tails of the distribution.

## RESULTS

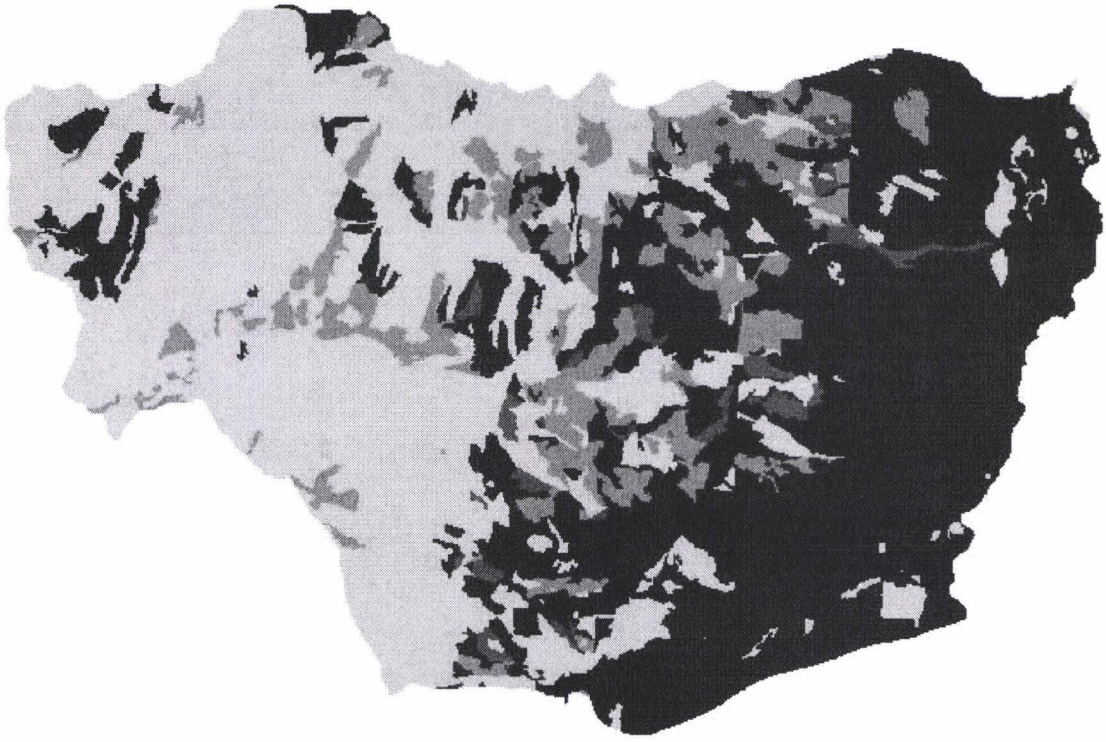
### Canopy Cover Landscape Metric (Shannon diversity index) SIMPPLE Simulations for Decade 10 (n = 100)

Histogram

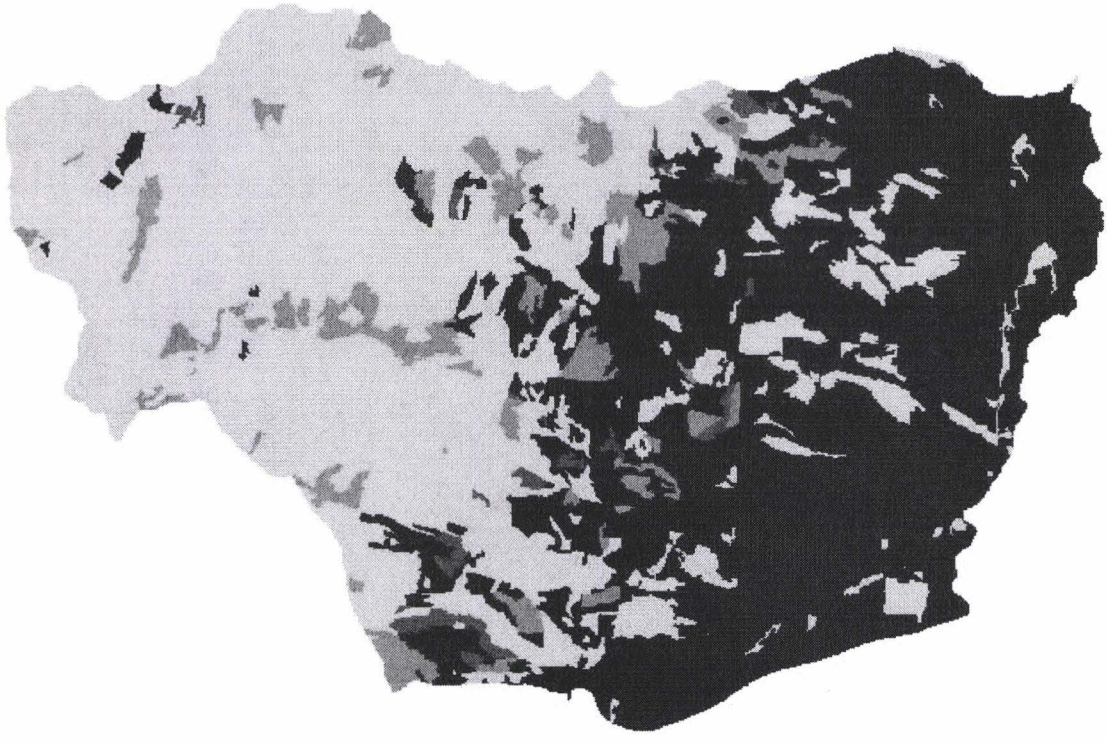
SHDI



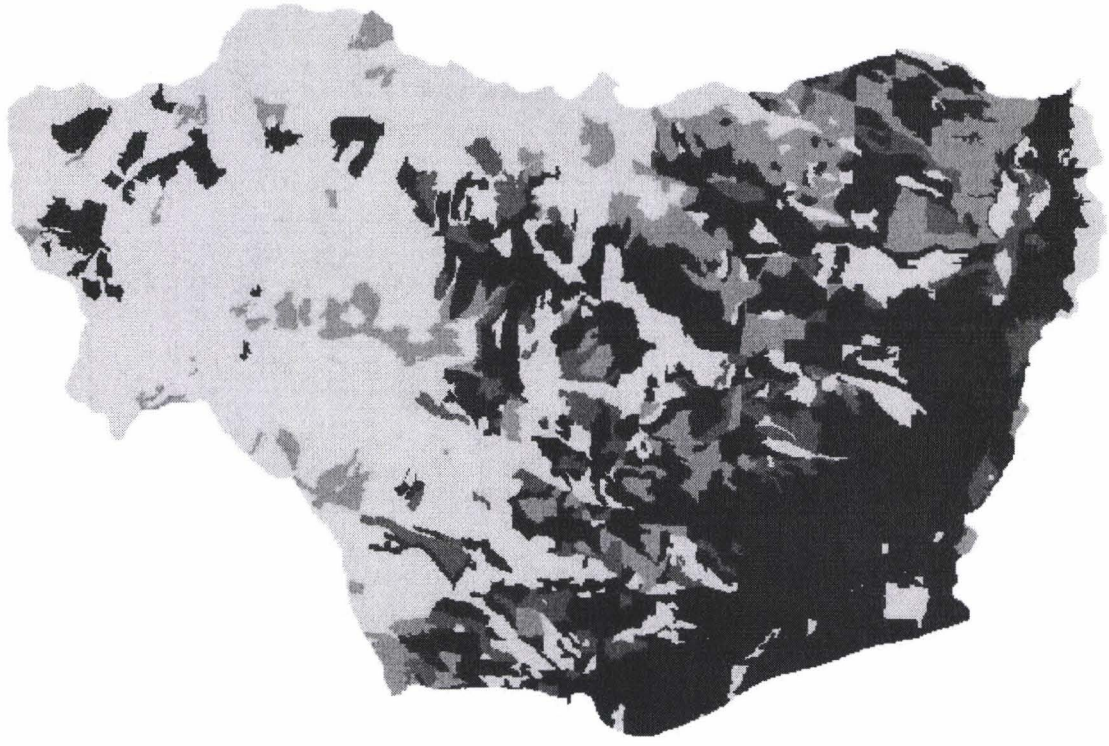
Mean landscape (SHDI = 1.031418)



Landscape at negative two standard deviations



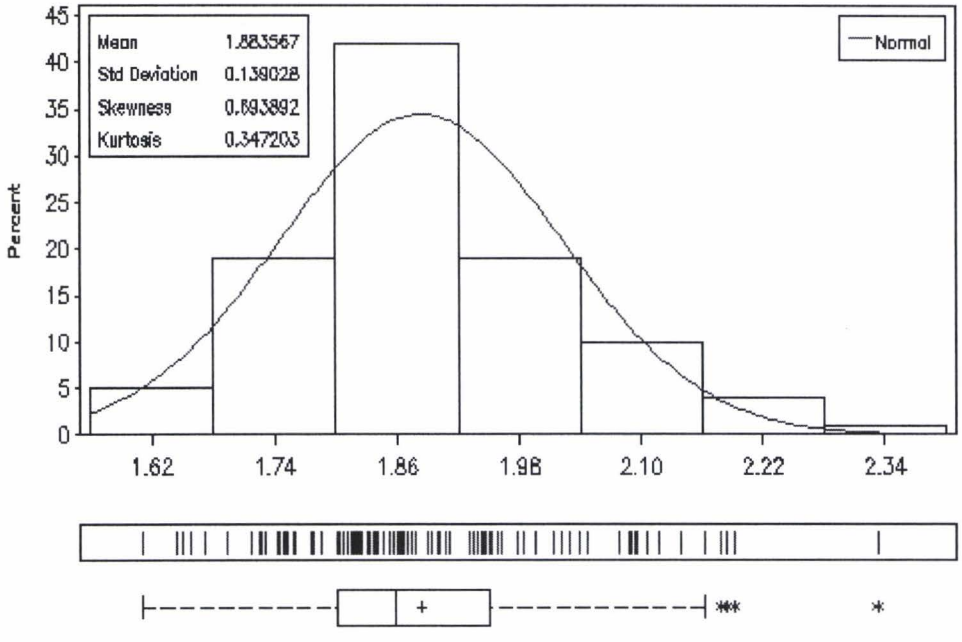
Landscape at positive two standard deviations



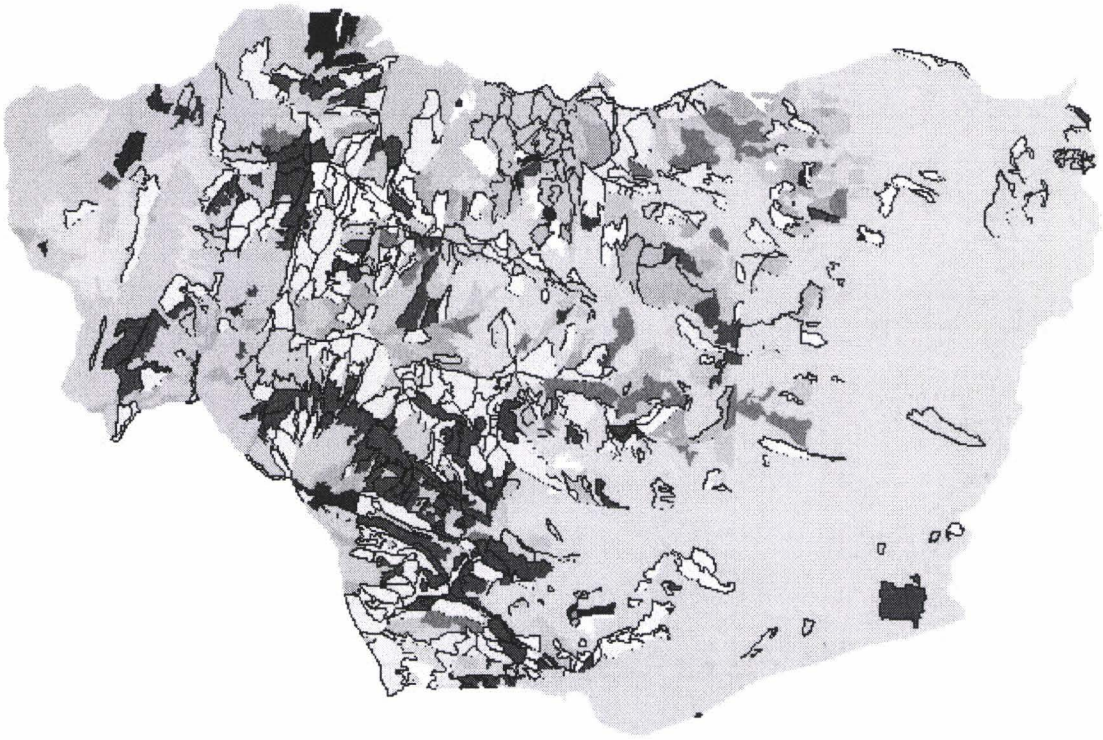
**Size Landscape Metric (Shannon diversity index)**
**SIMPPLE Simulations for Decade 10 (n = 100)**

Histogram

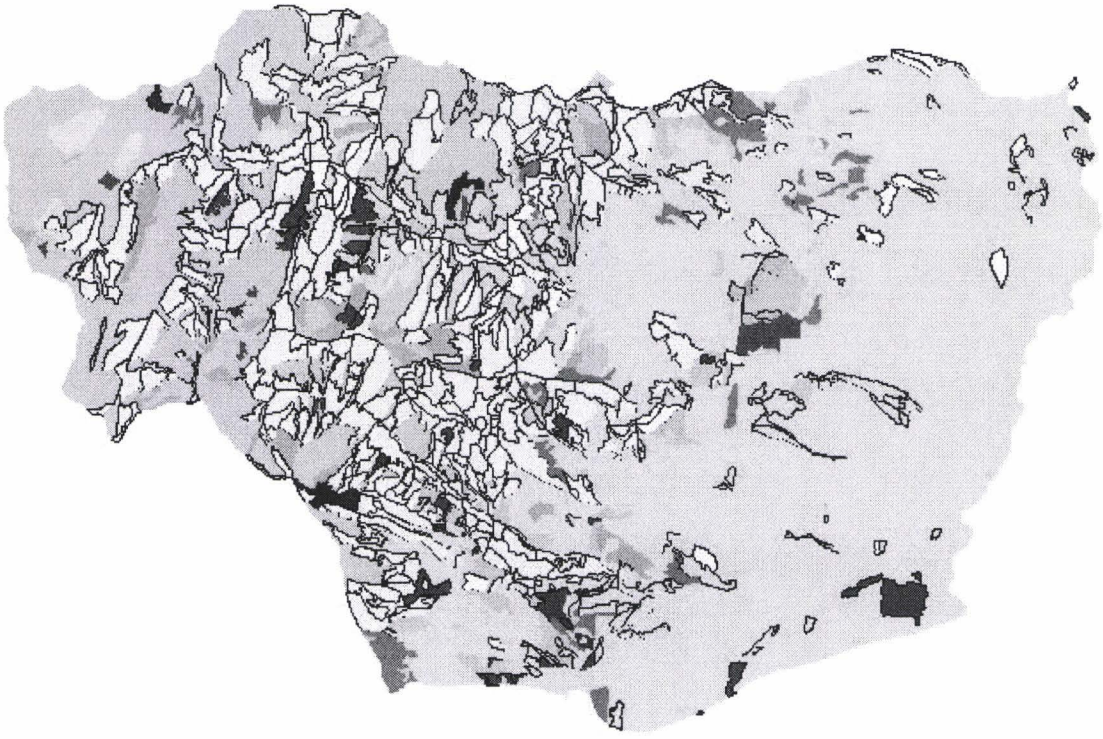
SHDI



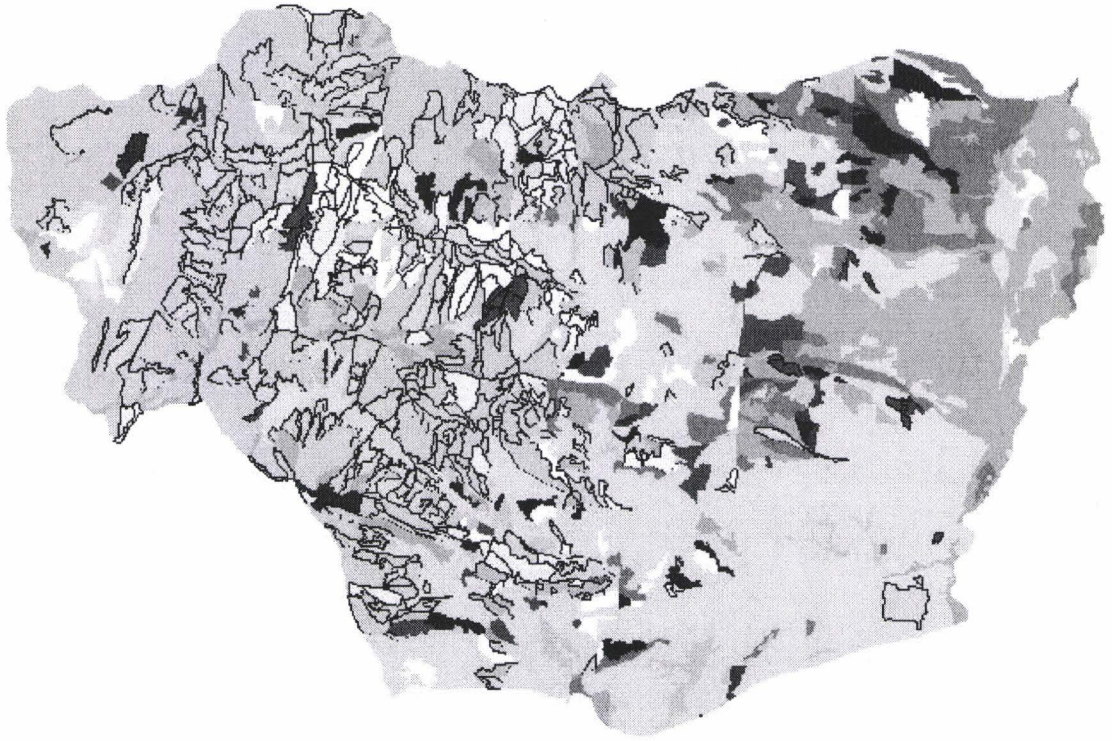
Mean landscape (SHDI = 1.883567)



Landscape at negative two standard deviations

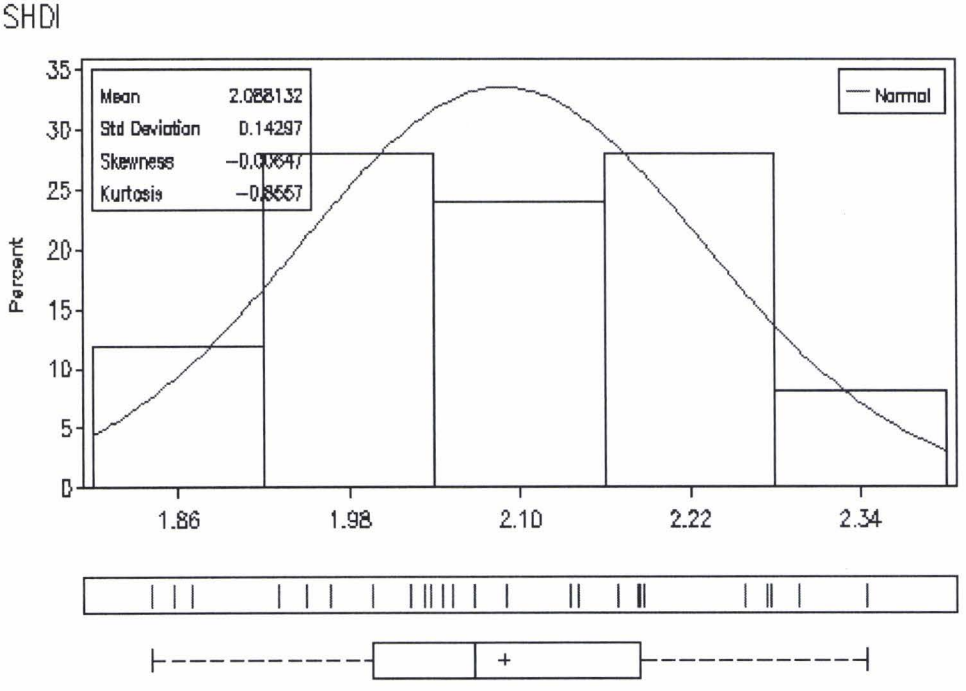


Landscape at positive two standard deviations

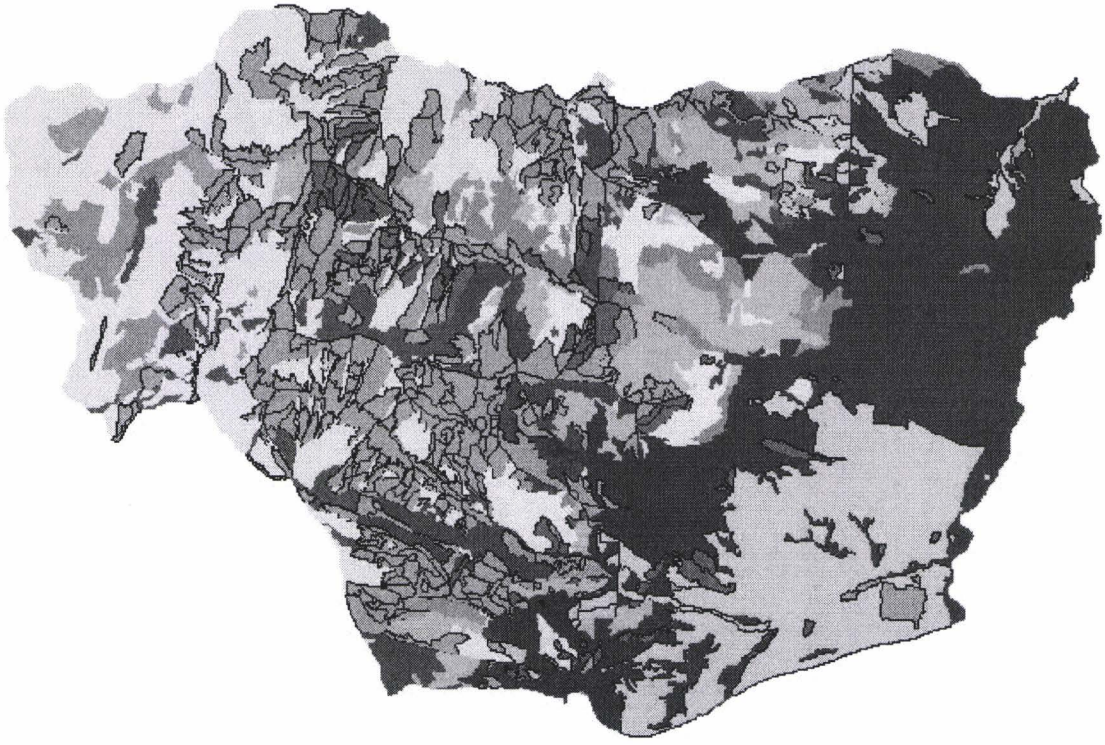


**Species Landscape Metric (Shannon diversity index)**
**SIMPPLE Simulations for Decade 10 (n = 100)**

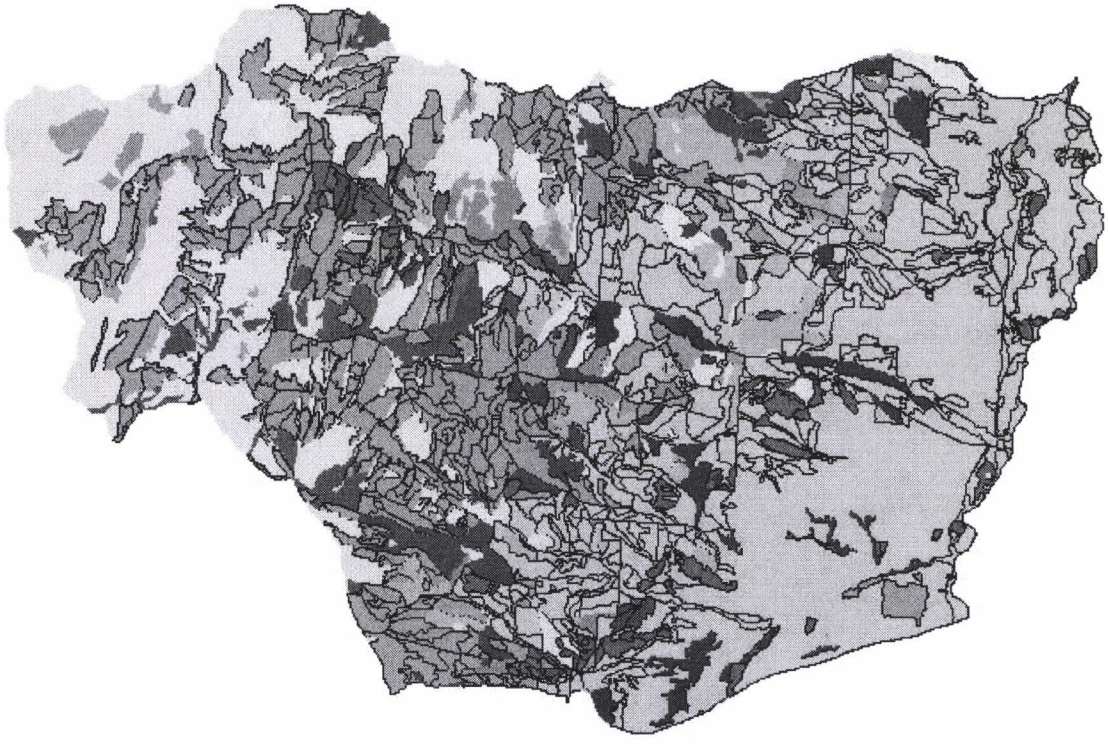
Histogram



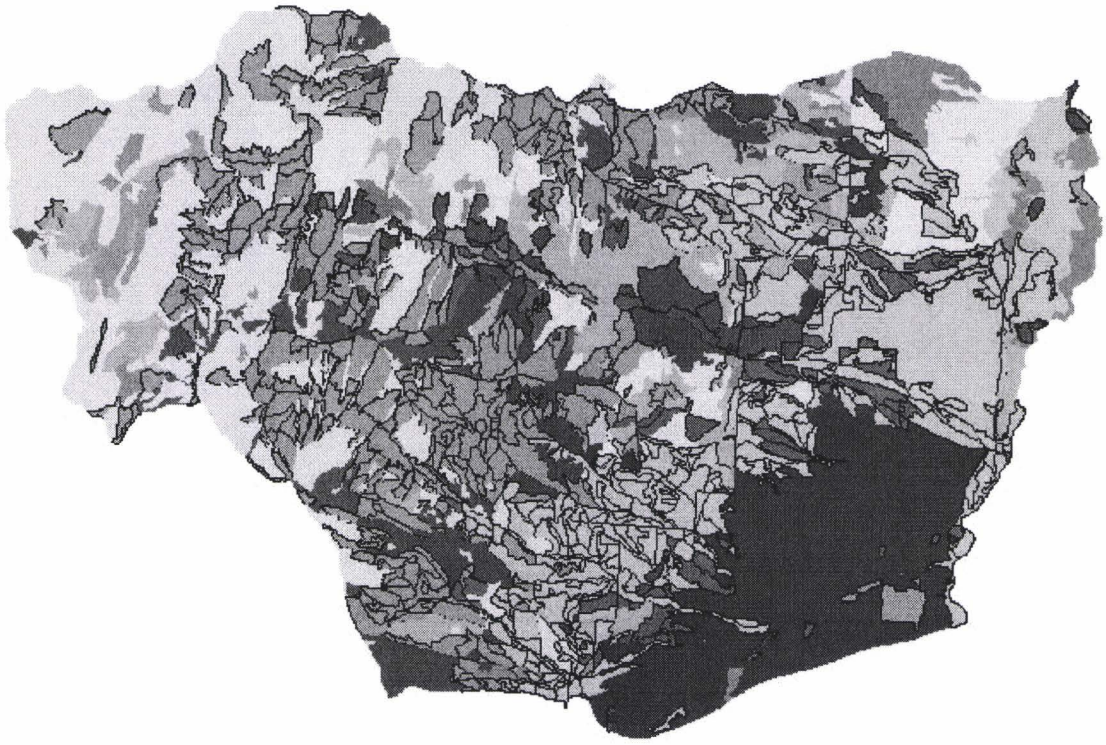
Mean landscape (SHDI = 2.088132)



Landscape at negative two standard deviations



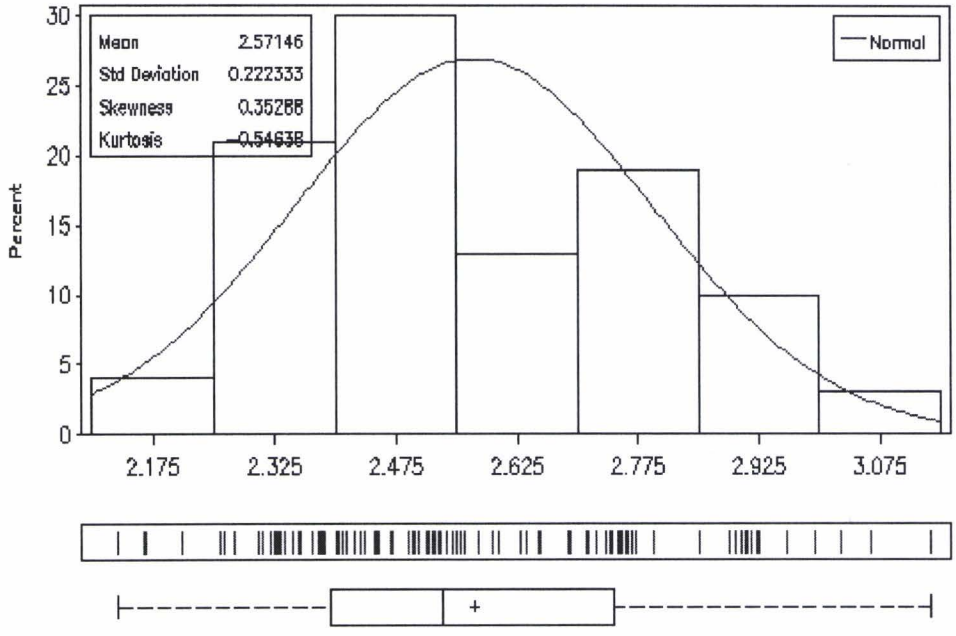
Landscape at positive two standard deviations



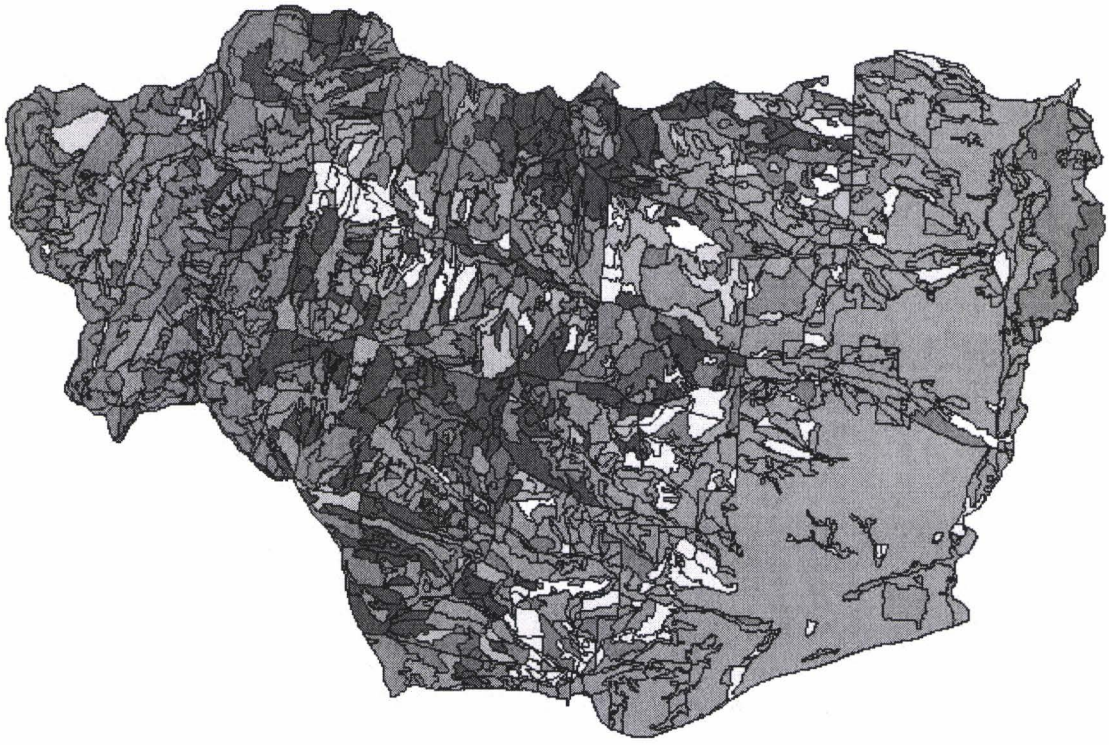
**State Landscape Metric (Shannon diversity index)**
**SIMPPLE Simulations for Decade 10 (n = 100)**

Histogram

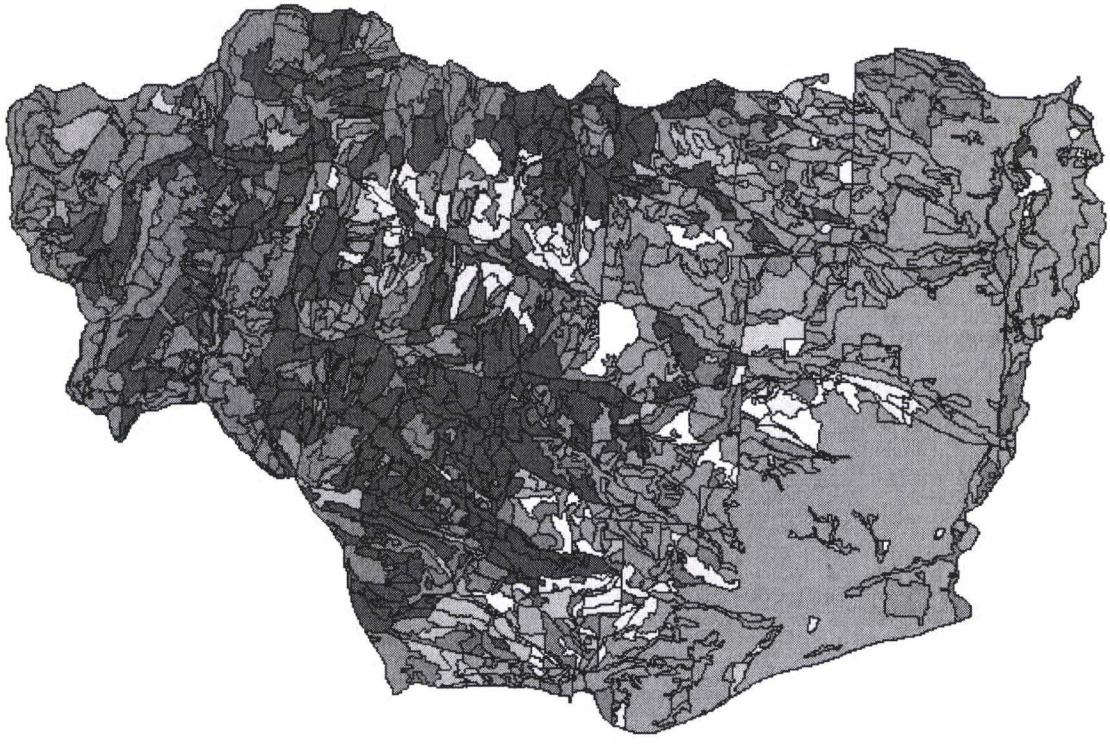
SHDI



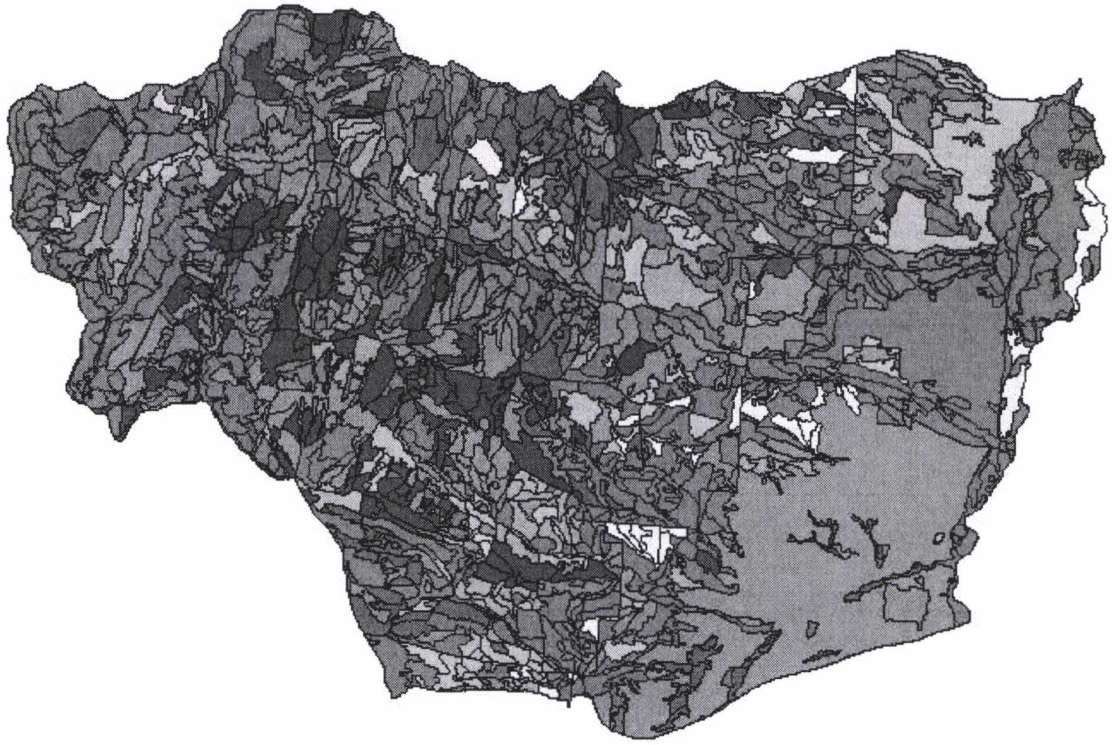
Mean landscape (SHDI = 2.57146)



Landscape at negative two standard deviations

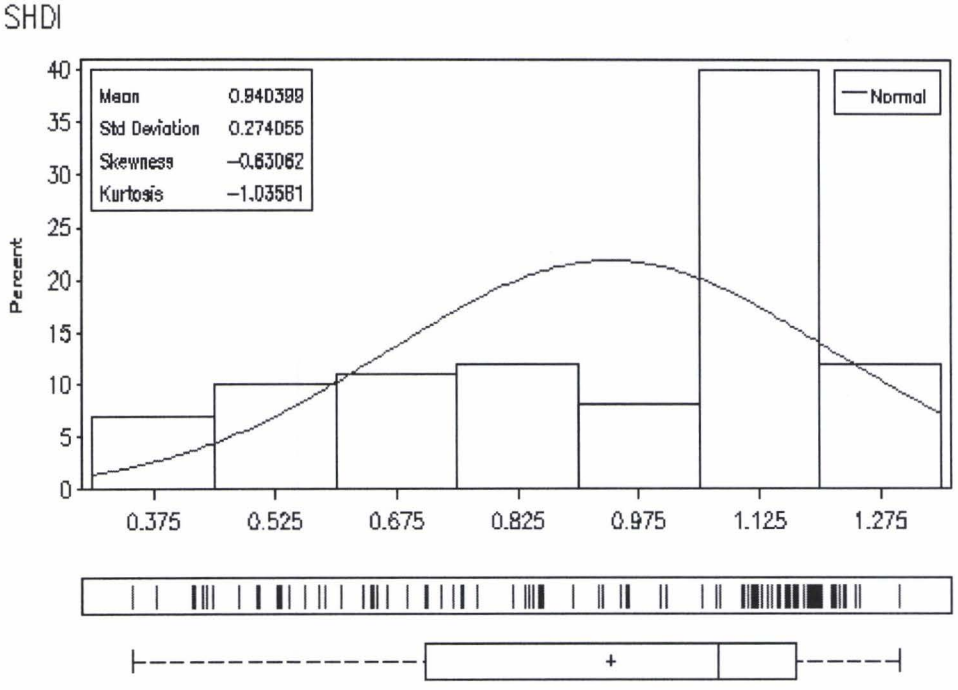


Landscape at positive two standard deviations

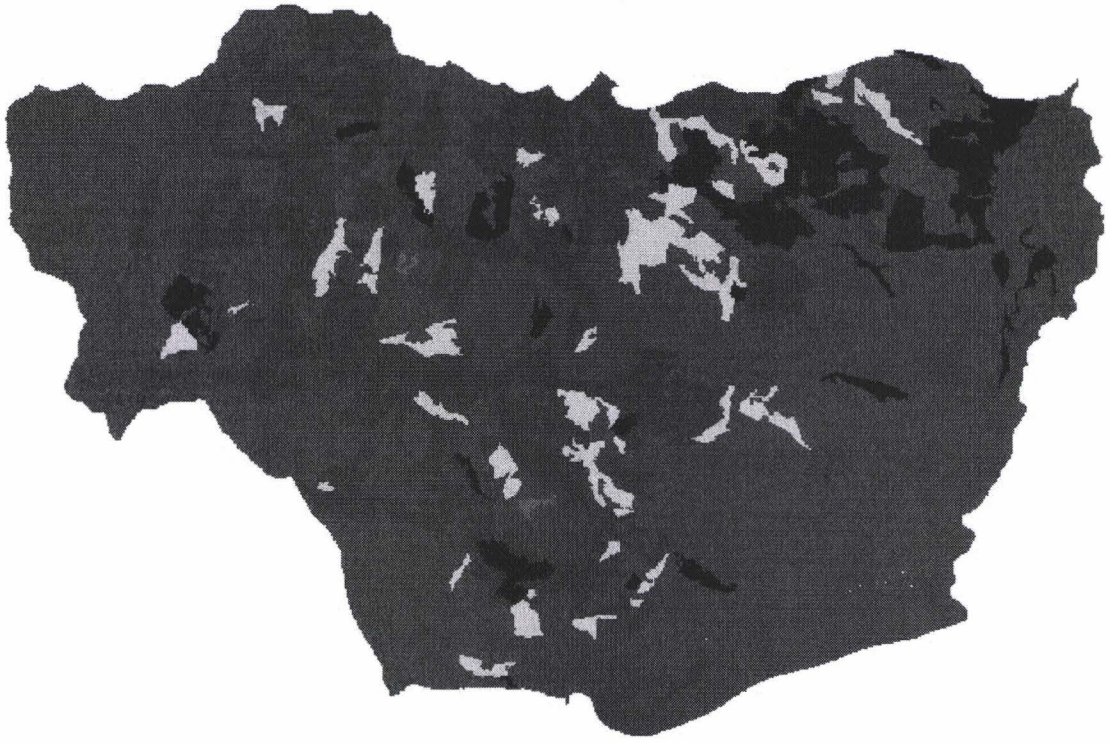


**Process Landscape Metric (Shannon diversity index)**
**SIMPPLE Simulations for Decade 10 (n = 100)**

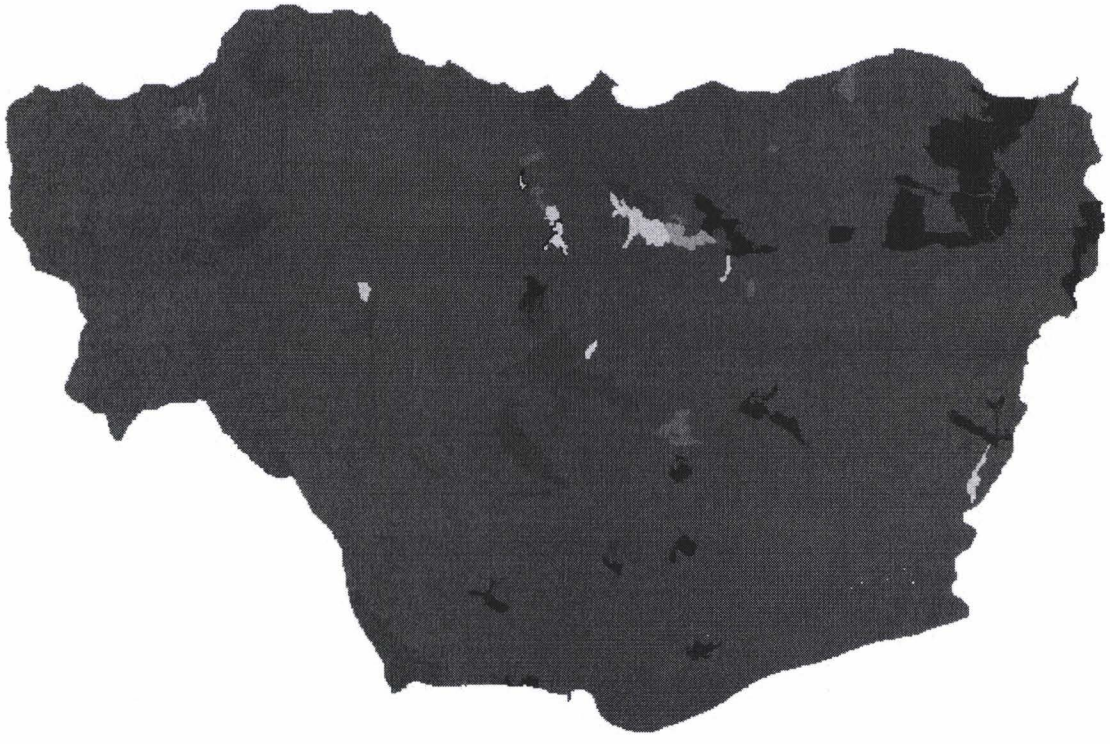
Histogram



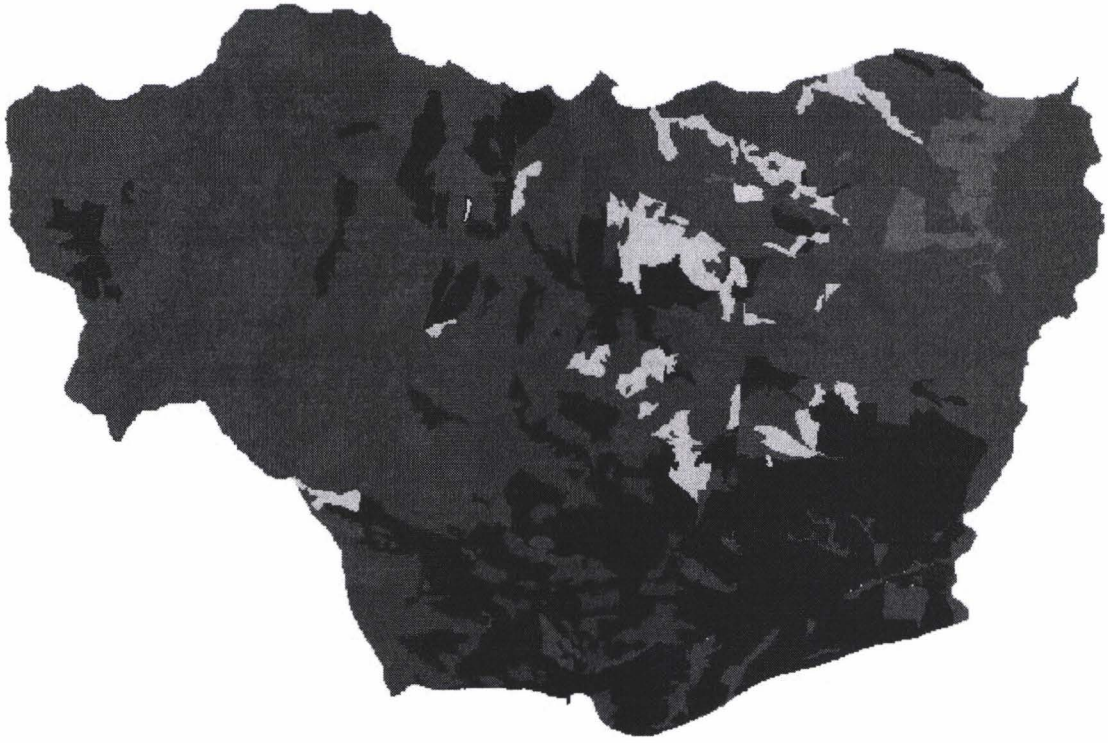
Mean landscape (SHDI = 0.940399)



Landscape at negative two standard deviations



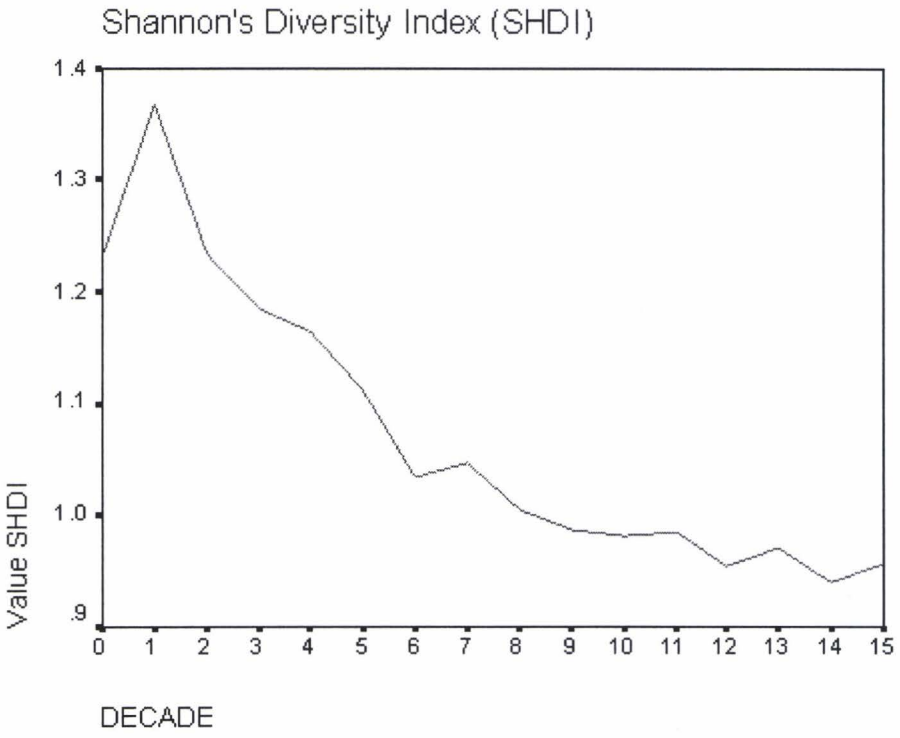
Landscape at positive two standard deviations



The previous pages presented results based on 100 independent simulations. Our investigation also considered the evaluation of these landscape metrics within a simulation. We present a display of SHDI values over fifteen decades for one simulation only. Observe that the SHDI values decrease steadily after decade 1 and that these observations are not independent. Calculating SHDI differences between decades may be one mechanism for restoring independence, but the stand attributes are nominal values and mathematical operations cannot be performed. The conditions under which landscape metrics can be applied within a single multi-decade simulation is an area of research that needs further investigation. As food-for-thought on this problem, displays are presented for the overall trend across all fifteen decades, and map displays for the minimum, mean, and maximum SHDI values for canopy cover and process attributes.

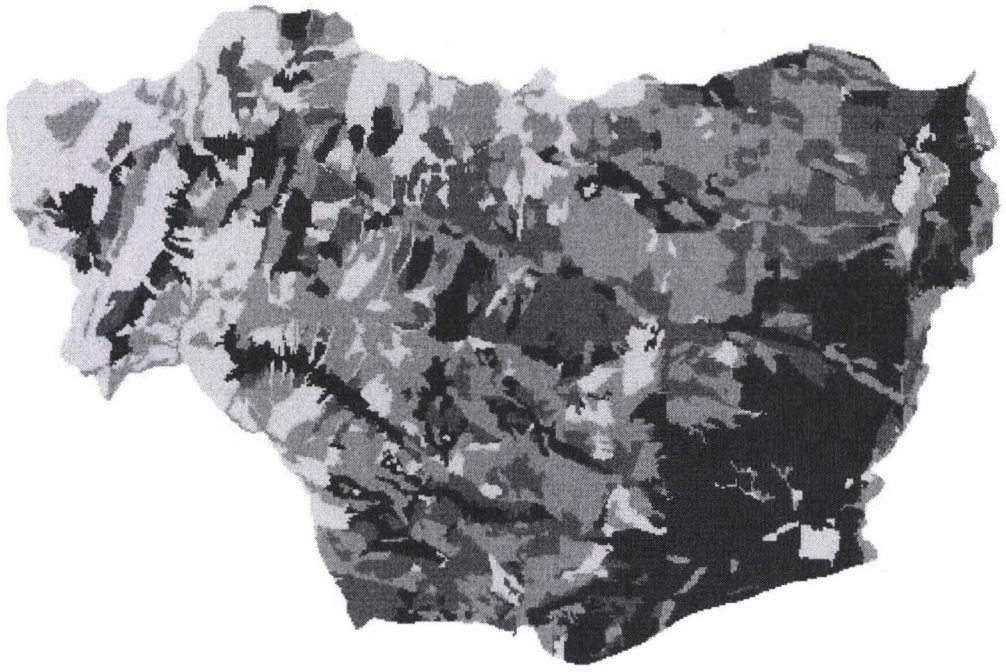
## Canopy Cover Landscape Metric (Shannon diversity index)

SIMPPLLE Simulation #61, Decades 0 to 15

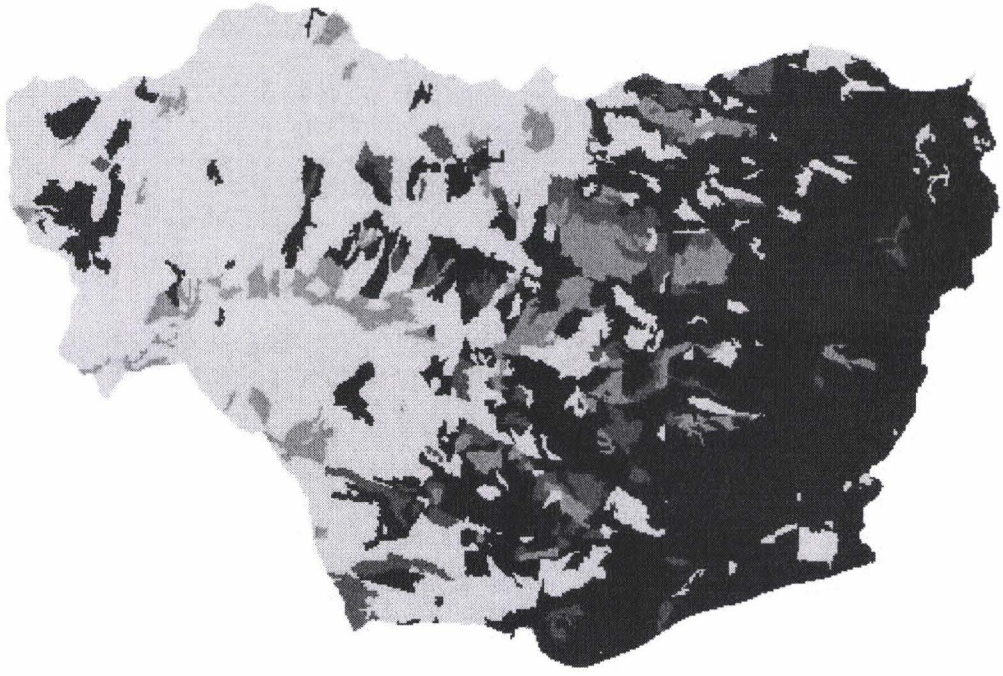



Mean SHDI value: 1.07, Standard deviation: 0.12

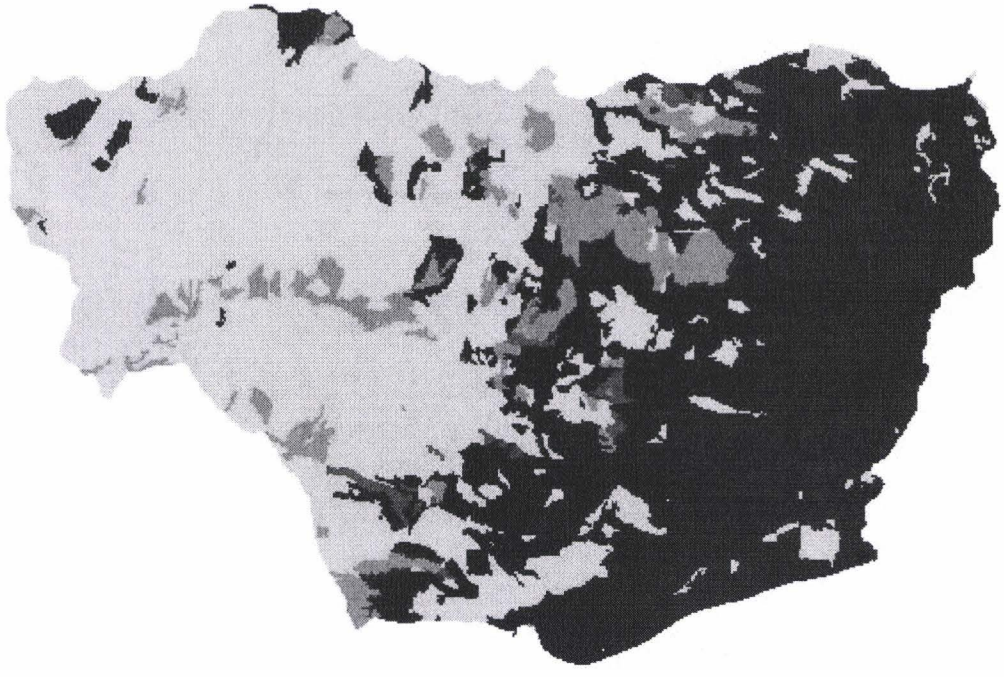
Canopy Cover SHDI for Decade 1 (minimum SHDI for simulation)



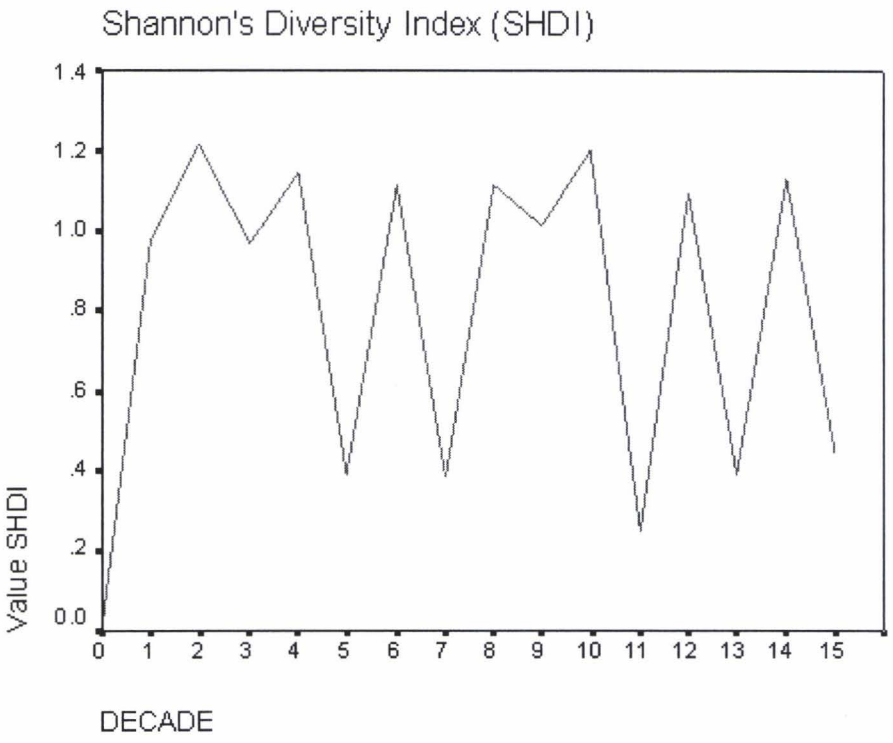
Canopy Cover SHDI for Decade 7 (mean SHDI for simulation)



Canopy Cover SHDI for Decade 14 (maximum SHDI for simulation)

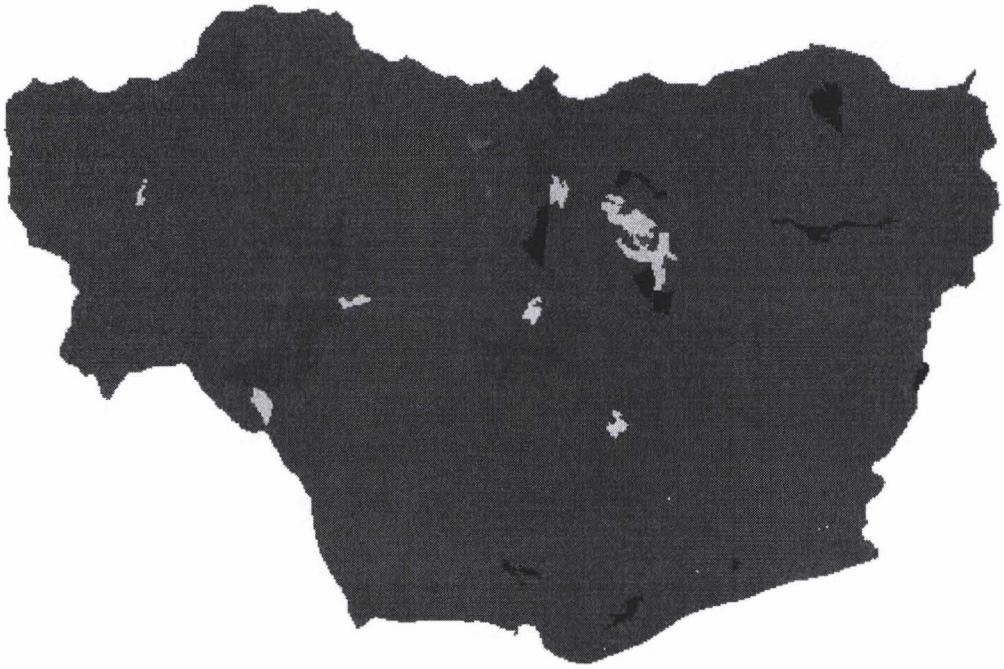


**Process Landscape Metric (Shannon diversity index)**
SIMPPLLE Simulation #61, Decades 0 to 15

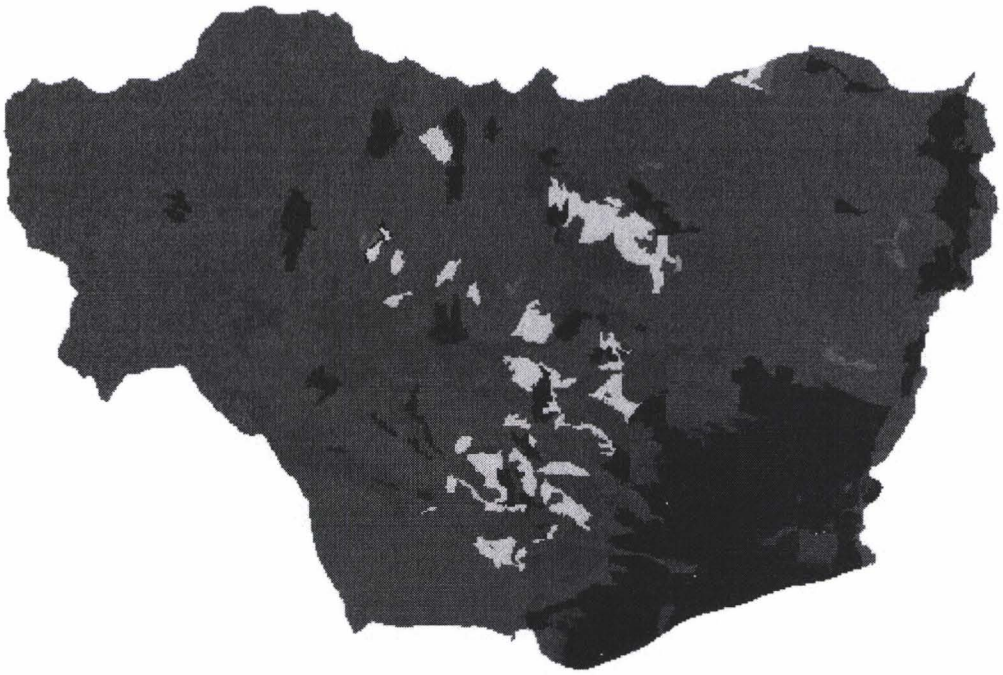



Mean SHDI value: 0.80, Standard deviation: 0.40

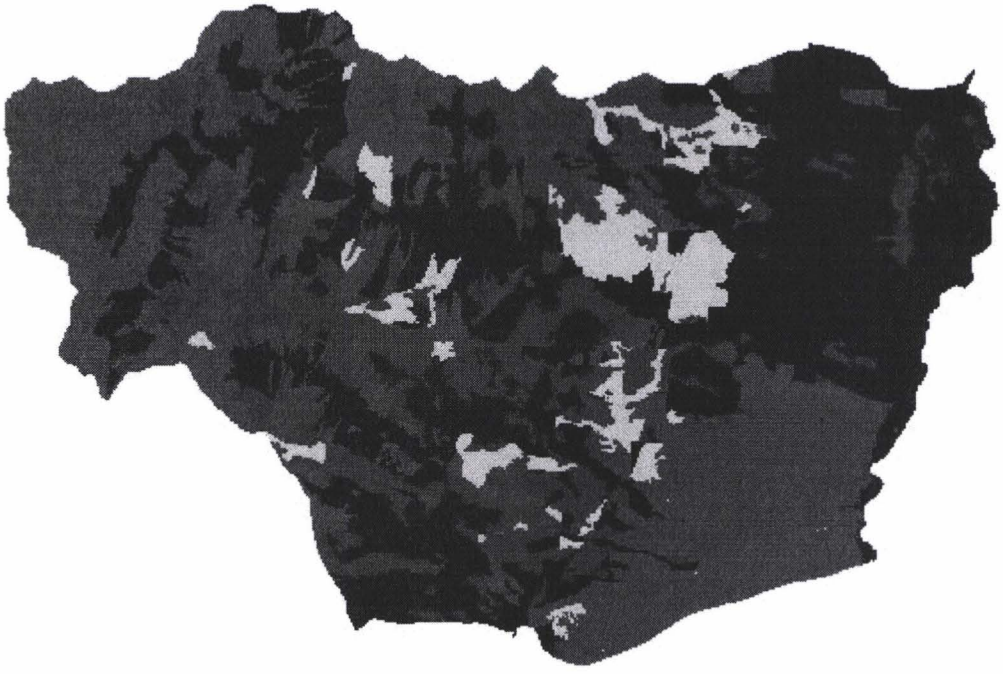
Process SHDI for Decade 11 (minimum SHDI for simulation)



Process SHDI for Decade 3 (mean SHDI for simulation)



Process SHDI for Decade 2 (maximum SHDI for simulation)



## Recommendations

Based on our investigation, the following recommendations can be made:

1. The landscape metrics investigated in this study follow the statistical assumptions of a normal distribution for independent simulations of a landscape within SIMPPLLE. The calculation of these landscape metrics and associated statistics could be incorporated within SIMPPLLE when the user selects multiple simulations, given that the number of simulations is large enough to guarantee that the result is normally distributed. A test for normality on the metrics could guide the user in making this determination.
2. User support in the form of help pages would have to be added to provide the user with some guidance on the different metrics and their correct application and interpretation.
3. Enhancements would have to be incorporated into the SIMPPLLE interface to provide for an effective and efficient display of the calculated metrics. At a minimum this would include histograms and basic statistics for the results. The addition of standardized legends for polygon (state) attributes would greatly assist the user in developing visual displays.
4. When landscape metrics were being computed we found that five significant digits to the right of the decimal were required to accurately summarize their values.
5. On hindsight we conclude that the FRAGSTATS module provided no advantage over writing AMLs to compute landscape metrics.

### Potential Future Work

Further investigation is required to:

1. Determine how these landscape metrics and the associated statistics might be applied to evaluate model performance. For example, an investigator may have certain expectations (a null hypothesis) or a test data set to evaluate model performance against. The metrics presented here may offer a statistical test to evaluate whether or not a specific scenario is within or outside a given distribution.
2. Determine the appropriateness of these metrics for interpretations within a single simulation where observations are not independent (i.e. decade 11 is dependent on decade 10).
3. Determine the value of these metrics to landscape interpretations within the context of SIMPPLLE.
4. Develop visual displays that provide for the efficient and effective interpretation of results.
5. Determine how these metrics and statistics help in the evaluation of two or more comparative landscapes (e.g. two watersheds within the same ecological zone)
6. Address landscape configuration since any two simulations could produce equivalent metrics, but the distribution within the landscape could be totally different.

## LITERATURE CITED

McGarigal, Kevin, and Barbara J. Marks 1994. *Fragstats: Spatial Pattern Analysis Program for Quantifying Landscape Structure*, Oregon State University, Corvallis, OR. 67pp.

Sokal, Robert R., and F. James Rohlf 1081. *Biometry,* W.H. Freeman and Company, San Francisco. 859 pp.

Zar, Jerold H. 1996. *Biostatistical Analysis*. 3rd ed. Prentice Hall, Upper Saddle River, New Jersey. 662 pp.